

[illegible]

Surendra Goel

America Online, Inc.

Searching Content on Web Pages

Fish & Richardson P.C.
601 Thirteenth Street, NW
Washington, DC 20005
Tel.: (202) 783-5070
Fax: (202) 783-2331

06975-152001

Searching Content on Web Pages

This application claims priority from U.S. Provisional Application No. 60/223,695, filed August 8, 2000, which is incorporated by reference.

TECHNICAL FIELD

This invention relates to searching systems and processes, and more particularly to searching content on web pages.

BACKGROUND

With the explosion of information on the Internet, it has become increasingly difficult to conduct a search on the Internet that returns results in a manner and a format that are useful to the person conducting the search. Frequently, when a search is performed, the most useful and relevant results may be scattered and buried among thousands of results.

In other instances, when a search is performed on the Internet, a search may yield few or no results even though relevant results exist on the Internet. Few or no results may occur because the Internet sites and the web pages within Internet sites that contain the desired results may not be searchable.

SUMMARY

In one general aspect, performing a search to identify web sites that relate to a search term based on text within the web sites includes receiving at least one search term that then is compared with electronic information within at least one electronic information store to determine whether matches exist. The electronic information within the at least one information store includes text displayed by different web pages from different web sites. Results based on the matches that are determined to exist are displayed. The results include at least one website identifier.

Embodiments may include one or more of the following features. For example, several search terms may be received and grouped as a single string by default.

The electronic information within the electronic information store also may include titles, descriptions, and addresses of web sites. Additionally, the electronic information within the electronic information store may include full text displayed by different web pages

from different web sites. The electronic information also may include text and/or full text of an introductory page displayed by different web pages from different web sites.

The search may be performed by a web host having members and may further include automatically scanning and storing the text of a website when the website is accessed by a member of the web host. The stored text may be compared against received search terms. The search may further include determining whether the text of the website accessed by a member has been previously stored.

Automatic scanning and storing of the text of the website being accessed by a member of the web host may occur when the text is determined not to have been previously stored. The determining may be based on a website address corresponding to the website. Additionally or alternatively, the determining may be based on the text of the website.

Web sites provided by a listing service may be identified, and the text of the web sites provided by the listing service that were not stored previously may be automatically scanned and stored. Identifying the web sites provided by the listing service may be performed periodically.

The full text of a website may be automatically scanned when the website is accessed by a member of the web host. The full text of at least a website provided by the listing service that has not been accessed by a member of the web host also may be automatically scanned.

The displayed results may include identifiers for several web sites. The identifiers may be ranked based on a number of matches that are determined to exist between the search term and the electronic information. Results may be communicated based on the matches that are determined to exist. The results may include at least one website identifier.

The ranking of the identifiers for the several web sites may be based on whether the matches occur within the text, the titles, the descriptions, or the addresses of the web sites. Matches that occur within more than one of the text, the titles, the descriptions, and the addresses of the website may be ranked higher than the results that include matches that occur within only one of the text, the titles, the descriptions, and the addresses of the website.

In another general aspect, performing a search to identify web sites that relate to a search term may include receiving at least one search term that then is compared with a list of recommended web sites, previously performed searches, and with electronic information

within at least one electronic information store to determine whether matches exist. The electronic information within the electronic information store may include text displayed by different web pages from different web sites. Results based on the matches that are determined to exist then are displayed.

5 Embodiments may include one or more of the following features. For example, the electronic information within the electronic information store may include full text displayed by different web pages from different web sites.

10 In another general aspect, a web host having members may populate at least one memory store by automatically scanning text of a website when the website is accessed by a member of the web host and storing the text of the website that was automatically scanned for comparison against search terms that are received.

15 In another general aspect, storing searchable content may include using first and second electronic regions that include text displayed by different web pages from different web sites. The first electronic region is populated by automatically scanning and storing the text of a website when the website is accessed a threshold number of times by members of a web host. The second electronic region is populated by automatically scanning and storing the text of a website provided by a listing service that was not accessed the threshold number of times by members of the web host.

20 Embodiments may include one or more of the following features. For example, the first electronic region and the second electronic region may include the full text, titles, descriptions, and addresses displayed by different web pages from different web sites. The text may include a symbol other than an alphanumeric symbol.

25 These general and specific aspects may be implemented using a system, a method, or a computer program, or any combination of systems, methods, and computer programs.

 Other features and advantages will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

Fig. 1 is a block diagram of a communications system.

Figs. 2-6 are expansions of the block diagram of Fig. 1.

Fig. 7 is a flowchart of a process for performing an electronic search, which may be performed by components of the systems shown in Figs. 1-6.

Fig. 8a is a flowchart of a process for performing a category search as part of the process of Fig. 7.

Fig. 8b is an exemplary screen shot that shows the results of the category search performed in Fig. 8a.

Fig. 8c is an exemplary screen shot that shows more detailed results of the category search performed in Fig. 8a.

Fig. 9a is a flowchart of a process for performing a web site search as part of the process of Fig. 7.

Fig. 9b is a flowchart of a process for searching different data stores as part of the process of Fig. 9a.

Fig. 9c is a block diagram of a system for storing searchable content.

Fig. 9d is a flowchart of a process for populating electronic information stores.

Fig. 9e is an exemplary screen shot that shows the results of the web site search.

Fig. 10a is a flowchart of a process for displaying search results as part of the process of Fig. 7.

Fig. 10b is an exemplary screen shot that shows the displayed search results.

Fig. 10c is an exemplary screen shot that shows matching web page results.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

For illustrative purposes, Figs. 1-6 describe a communications system for implementing techniques for transferring files between subscribers of an instant messaging host complex. For brevity, several elements in the figures described below are represented as monolithic entities. However, as would be understood by one skilled in the art, these elements each may include numerous interconnected computers and components designed to perform a set of specified operations and/or dedicated to a particular geographical region.

Referring to Fig. 1, a communications system 100 is capable of delivering and exchanging data between a client system 105 and a host system 110 through a communications link 115. The client system 105 typically includes one or more client

devices 120 and/or client controllers 125, and the host system 110 typically includes one or more host devices 135 and/or host controllers 140. For example, the client system 105 or the host system 110 may include one or more general-purpose computers (e.g., personal computers), one or more special-purpose computers (e.g., devices specifically programmed to communicate with each other and/or the client system 105 or the host system 110), or a combination of one or more general-purpose computers and one or more special-purpose computers. The client system 105 and the host system 110 may be arranged to operate within or in concert with one or more other systems, such as, for example, one or more LANs ("Local Area Networks") and/or one or more WANs ("Wide Area Networks").

The client device 120 (or the host controller 135) is generally capable of executing instructions under the command of a client controller 125 (or a host controller 140). The client device 120 (or the host device 135) is connected to the client controller 125 (or the host controller 140) by a wired or wireless data pathway 130 or 145 capable of delivering data.

The client device 120, the client controller 125, the host device 135, and the host controller 140 each typically include one or more hardware components and/or software components. An example of a client device 120 or a host device 135 is a general-purpose computer (e.g., a personal computer) capable of responding to and executing instructions in a defined manner. Other examples include a special-purpose computer, a workstation, a server, a device, a component, other physical or virtual equipment or some combination thereof capable of responding to and executing instructions.

An example of client controller 125 or a host controller 140 is a software application loaded on the client device 120 or the host device 135 for commanding and directing communications enabled by the client device 120 or the host device 135. Other examples include a program, a piece of code, an instruction, a device, a computer, a computer system, or a combination thereof, for independently or collectively instructing the client device 120 or the host device 135 to interact and operate as described. The client controller 125 and the host controller 140 may be embodied permanently or temporarily in any type of machine, component, physical or virtual equipment, storage medium, or propagated signal capable of providing instructions to the client device 120 or the host device 135.

The communications link 115 typically includes a delivery network 160 making a direct or indirect communication between the client system 105 and the host system 110,

irrespective of physical separation. Examples of a delivery network 160 include the Internet, the World Wide Web, WANs, LANs, analog or digital wired and wireless telephone networks (e.g. PSTN, ISDN, and xDSL), radio, television, cable, satellite, and/ or any other delivery mechanism for carrying data. The communications link 115 may include communication pathways 150, 155 that enable communications through the one or more delivery networks 160 described above. Each of the communication pathways 150, 155 may include, for example, a wired, wireless, cable or satellite communication pathway.

Fig. 2 illustrates a communication system 200 including a client system 205 communicating with a host system 210 through a communications link 215. Client system 205 typically includes one or more client devices 220 and one or more client controllers 225 for controlling the client devices 220. Host system 210 typically includes one or more host devices 235 and one or more host controllers 240 for controlling the host devices 235. The communications link 215 may include communication pathways 250, 255 enabling communications through the one or more delivery networks 260.

Examples of each element within the communication system of Fig. 2 are broadly described above with respect to Fig. 1. In particular, the host system 210 and the communications link 215 typically have attributes comparable to those described with respect to the host system 110 and the communications link 115 of Fig. 1, respectively. Likewise, the client system 205 of Fig. 2 typically has attributes comparable to and may illustrate one possible embodiment of the client system 105 of Fig. 1.

The client device 220 typically includes a general purpose computer 270 having an internal or external storage 272 for storing data and programs such as an operating system 274 (e.g., DOS, Windows™, Windows 95™, Windows 98™, Windows 2000™, Windows NT™, OS/2, and Linux) and one or more application programs. Examples of application programs include authoring applications 276 (e.g., word processing, database programs, spreadsheet programs, and graphics programs) capable of generating documents or other electronic content; client applications 278 (e.g., AOL client, CompuServe client, AIM client, AOL TV client, and ISP client) capable of communicating with other computer users, accessing various computer resources, and viewing, creating, or otherwise manipulating electronic content; and browser applications 280 (e.g., Netscape's Navigator and Microsoft's Internet Explorer) capable of rendering standard Internet content.

The general-purpose computer 270 also includes a central processing unit 282 (CPU) for executing instructions in response to commands from the client controller 225. In one implementation, the client controller 225 includes one or more of the application programs installed on the internal or external storage 272 of the general-purpose computer 270. In
5 another implementation, the client controller 225 includes application programs externally stored in and executed by one or more device(s) external to the general- purpose computer 270.

The general-purpose computer typically will include a communication device 284 for sending and receiving data. One example of the communication device 284 is a modem.
10 Other examples include a transceiver, a set-top box, a communication card, a satellite dish, an antenna, or another network adapter capable of transmitting and receiving data over the communications link 215 through a wired or wireless data pathway 250. The general-purpose computer 270 also may include a TV ("television") tuner 286 for receiving television programming in the form of broadcast, satellite, and/or cable TV signals. As a result, the
15 client device 220 can selectively and/or simultaneously display network content received by communications device 284 and television programming content received by the TV tuner 286.

The general-purpose computer 270 typically will include an input/output interface 288 to enable a wired or wireless connection to various peripheral devices 290. Examples of
20 peripheral devices 290 include, but are not limited to, a mouse 291, a mobile phone 292, a personal digital assistant 293 (PDA), a keyboard 294, a display monitor 295 with or without a touch screen input, and/or a TV remote control 296 for receiving information from and rendering information to subscribers. Other examples may include voice recognition and synthesis devices.

Although Fig. 2 illustrates devices such as a mobile telephone 292, a PDA 293, and a
25 TV remote control 296 as being peripheral with respect to the general-purpose computer 270, in another implementation, such devices may themselves include the functionality of the general-purpose computer 270 and operate as the client device 220. For example, the mobile phone 292 or the PDA 293 may include computing and networking capabilities, and may
30 function as a client device 220 by accessing the delivery network 260 and communicating

with the host system 210. Furthermore, the client system 205 may include one, some or all of the components and devices described above.

Referring to Fig. 3, a communications system 300 is capable of delivering and exchanging information between a client system 305 and a host system 310 through a communication link 315. Client system 305 typically includes one or more client devices 320 and one or more client controllers 325 for controlling the client devices 320. Host system 310 typically includes one or more host devices 335 and one or more host controllers 340 for controlling the host devices 335. The communications link 315 may include communication pathways 350, 355 enabling communications through the one or more delivery networks 360.

Examples of each element within the communication system of Fig. 3 are broadly described above with respect to Figs. 1 and 2. In particular, the client system 305 and the communications link 315 typically have attributes comparable to those described with respect to client systems 105 and 205 and communications links 115 and 215 of Figs. 1 and 2. Likewise, the host system 310 of Fig. 3 may have attributes comparable to and may illustrate one possible embodiment of the host systems 110 and 210 shown in Figs. 1 and 2.

The host system 310 includes a host device 335 and a host controller 340. The host controller 340 is generally capable of transmitting instructions to any or all of the elements of the host device 335. For example, in one implementation, the host controller 340 includes one or more software applications loaded on the host device 335. However, in other implementations, as described above, the host controller 340 may include any of several other programs, machines, and devices operating independently or collectively to control the host device 335.

The host device 335 includes a login server 370 for enabling access by subscribers and routing communications between the client system 305 and other elements of the host device 335. The host device 335 also includes various host complexes such as the depicted OSP ("Online Service Provider") host complex 380 and IM ("Instant Messaging") host complex 390. To enable access to these host complexes by subscribers, the client system 305 may include communication software, for example, an OSP client application and an IM client application. The OSP and IM communication software applications are designed to facilitate the subscriber's interactions with the respective services and, in particular, may

provide access to all the services available within the respective host complexes. For example, Instant Messaging allows a subscriber to use the IM client application to view whether particular subscribers ("buddies") are online, exchange instant messages with particular subscribers, participate in group chat rooms, trade files such as pictures, invitations or documents, find other subscribers with similar interests, get customized news and stock quotes, and search the Web.

Typically, the OSP host complex 380 supports different services, such as email, discussion groups, chat, news services, and Internet access. The OSP host complex 380 is generally designed with an architecture that enables the machines within the OSP host complex 380 to communicate with each other, certain protocols (i.e., standards, formats, conventions, rules, and structures) being employed to enable the transfer of data. The OSP host complex 380 ordinarily employs one or more OSP protocols and custom dialing engines to enable access by selected client applications. The OSP host complex 380 may define one or more specific protocols for each service based on a common, underlying proprietary protocol.

The IM host complex 390 is generally independent of the OSP host complex 380, and supports instant messaging services irrespective of a subscriber's network or Internet access. Thus, the IM host complex 390 allows subscribers to send and receive instant messages, whether or not they have access to any particular ISP. The IM host complex 390 may support associated services, such as administrative matters, advertising, directory services, chat, and interest groups related to the instant messaging. The IM host complex 390 has an architecture that enables all of the machines within the IM host complex to communicate with each other. To transfer data, the IM host complex 390 employs one or more standard or exclusive IM protocols.

The host device 335 may include one or more gateways that connect and therefore link complexes, such as the OSP host complex gateway 385 and the IM host complex gateway 395. The OSP host complex gateway 385 and the IM host complex 395 gateway may directly or indirectly link the OSP host complex 380 with the IM host complex 390 through a wired or wireless pathway. Ordinarily, when used to facilitate a link between complexes, the OSP host complex gateway 385 and the IM host complex gateway 395 are privy to information regarding a protocol anticipated by a destination complex, which

enables any necessary protocol conversion to be performed incident to the transfer of data from one complex to another. For instance, the OSP host complex 380 and IM host complex 390 may use different protocols such that transferring data between the complexes requires protocol conversion by or at the request of the OSP host complex gateway 385 and/or the IM host complex gateway 395.

Referring to Fig. 4, a communications system 400 is capable of delivering and exchanging information between a client system 405 and a host system 410 through a communication link 415. Client system 405 typically includes one or more client devices 420 and one or more client controllers 425 for controlling the client devices 420. Host system 410 typically includes one or more host devices 435 and one or more host controllers 440 for controlling the host devices 435. The communications link 415 may include communication pathways 450, 455 enabling communications through the one or more delivery networks 460. As shown, the client system 405 may access the Internet 465 through the host system 410.

Examples of each element within the communication system of Fig. 4 are broadly described above with respect to Figs. 1-3. In particular, the client system 405 and the communications link 415 typically have attributes comparable to those described with respect to client systems 105, 205, and 305 and communications links 115, 215, and 315 of Figs. 1-3. Likewise, the host system 410 of Fig. 4 may have attributes comparable to and may illustrate one possible embodiment of the host systems 110, 210, and 310 shown in Figs. 1-3. Fig. 4 describes an aspect of the host system 410, focusing primarily on one particular implementation of OSP host complex 480.

The client system 405 includes a client device 420 and a client controller 425. The client controller 425 is generally capable of establishing a connection to the host system 410, including the OSP host complex 480, the IM host complex 490 and/or the Internet 465. In one implementation, the client controller 425 includes an OSP application for communicating with servers in the OSP host complex 480 using OSP protocols that may or may not be exclusive or proprietary. The client controller 425 also may include applications, such as an IM client application and/or an Internet browser application, for communicating with the IM host complex 490 and the Internet 465.

The host system 410 includes a host device 435 and a host controller 440. The host controller 440 is generally capable of transmitting instructions to any or all of the elements of the host device 435. For example, in one implementation, the host controller 440 includes one or more software applications loaded on one or more elements of the host device 435. In other implementations, as described above, the host controller 440 may include any of several other programs, machines, and devices operating independently or collectively to control the host device 435.

The host device 435 includes a login server 470 capable of enabling communications between client systems 405 and various elements of the host system 410, including elements such as OSP host complex 480 and IM host complex 490. The login server 470 may implement one or more authorization procedures to enable simultaneous access to one or more of these elements.

The OSP host complex 480 and the IM host complex 490 are typically connected through one or more OSP host complex gateways 485 and one or more IM host complex gateways 495. Each OSP host complex gateway 485 and IM host complex gateway 495 may generally perform protocol conversions necessary to enable communication between one or more of the OSP host complex 480, the IM host complex 490, and the Internet 465.

The OSP host complex 480 supports a set of services to be accessed through and/or performed by from one or more servers located internal to and external from the OSP host complex 480. Servers external to the OSP host complex 480 may communicate using the Internet 465. Servers internal to the OSP complex 480 may be arranged in one or more configurations. For example, servers may be arranged in large centralized clusters identified as farms 4802 or in localized clusters identified as pods 4804.

More specifically, farms 4802 are groups of servers located at centralized locations within the OSP host complex 480. Farms 4802 generally are dedicated to providing particular functionality and services to subscribers and clients from a centralized location, regardless of the location of the subscriber or client. Farms 4802 are particularly useful for providing services that depend upon other remotely-located or performed processes and services for information, such as, for example, chat, email, instant messaging, news, newsgroups, search, stock updates, and weather. Thus, farms 4802 tend to rely on

connections with external resources such as the Internet 465 and/or other servers within the OSP host complex 480.

By contrast to farms 4802, pods 4804 are clusters of localized servers that provide some services offered by the OSP host complex 480 from a location local to the service or information recipient, which reduces and avoids time delays and congestion inherent in centralized processing. Each pod 4804 includes one or more interrelated servers capable of operating together to provide one or more services offered by the OSP host complex 480 in a geographically localized manner, with the servers of a pod 4804 generally operating independently of resources external to the pod 4804. A pod 4804 may cache content received from external sources, such as farms 4802 or the Internet 465, making frequently requested information readily available to the local service or information recipients served by the pod 4804. In this way, pods 4804 are particularly useful in providing services that are independent of other processes and servers such as, for example, routing to other localized resources or recipients, providing access to keywords and geographically specific content, providing access to routinely accessed information, and downloading certain software and graphical interface updates with reduced processing time and congestion. The determination of which servers and processes are located in the pod 4804 is made by the OSP according to load distribution, frequency of requests, demographics, and other factors.

In addition to farms 4802 and pods 4804, the implementation of Fig. 4 also includes one or more non-podded and non-farmed servers 4806. In general, the servers 4806 may be dedicated to performing a particular service or information that relies on other processes and services for information and may be directly or indirectly connected to resources outside of the OSP host complex 480, such as the Internet 465 and the IM host complex 490, through an OSP gateway 4808 within OSP host complex gateway 485. In the event that subscriber usage of a particular service or information of the servers 4806 becomes relatively high, those servers 4806 may be integrated into a farm or pod, as appropriate.

In the implementation of Fig. 4, one particular exemplary pod 4810 is shown in more detail. Pod 4810 includes a routing processor 4812. In a packet-based implementation, the client system 405 may generate information requests, convert the requests into data packets, sequence the data packets, perform error checking and other packet-switching techniques, and transmit the data packets to the routing processor 4812. Upon receiving data packets

from the client system 405, the routing processor 4812 may directly or indirectly route the data packets to a specified destination within or outside of the OSP host complex 480. In general, the routing processor 4812 will examine an address field of a data request, use a mapping table to determine the appropriate destination for the data request, and direct the data request to the appropriate destination.

For example, in the event that a data request from the client system 405 can be satisfied locally, the routing processor 4812 may direct the data request to a local server 4814 in the pod 4810. In the event that the data request cannot be satisfied locally, the routing processor 4812 may direct the data request internally to one or more farms 4802, one or more other pods 4804, or one or more non-podded servers 4806 in the OSP host complex 480, or the routing processor 4812 may direct the data request externally to elements such as the IM host complex 490 through an OSP/pod gateway 4816.

The routing processor 4812 also may direct data requests and/or otherwise facilitate communication between the client system 405 and the Internet 465 through the OSP/pod gateway 4816. In one implementation, the client system 405 uses an OSP client application to convert standard Internet content and protocols into OSP protocols and vice versa, where necessary. For example, when a browser application transmits a request in a standard Internet protocol, the OSP client application can intercept the request, convert the request into an OSP protocol and send the converted request to the routing processor 4812 in the OSP host complex 480. The routing processor 4812 recognizes the Internet 465 as the destination and routes the data packets to an IP ("Internet Protocol") tunnel 4818. The IP tunnel 4818 converts the data from the OSP protocol back into standard Internet protocol and transmits the data to the Internet 465. The IP tunnel 4818 also converts the data received from the Internet in the standard Internet protocol back into the OSP protocol and sends the data to the routing processor 4812 for delivery back to the client system 405. At the client system 405, the OSP client application converts the data in the OSP protocol back into standard Internet content for communication with the browser application.

The IP tunnel 4818 may act as a buffer between the client system 405 and the Internet 465, and may implement content filtering and time saving techniques. For example, the IP tunnel 4818 can check parental controls settings of the client system 405 and request and transmit content from the Internet 465 according to the parental control settings. In addition,

the IP tunnel 4818 may include a number a caches for storing frequently accessed information. If requested data is determined to be stored in the caches, the IP tunnel 4818 may send the information to the client system 405 from the caches and avoid the need to access the Internet 465.

5 In another implementation, the client system 405 may use standard Internet protocols and formatting to access pods 4810 and the Internet 465. For example, the subscriber can use an OSP TV client application having an embedded browser application installed on the client system 405 to generate a request in standard Internet protocol, such as HTTP ("HyperText Transport Protocol"). In a packet-based implementation, data packets may be encapsulated
10 inside a standard Internet tunneling protocol, such as, for example, UDP ("User Datagram Protocol"), and routed to a web tunnel 4820. The web tunnel 4820 may be a L2TP ("Layer Two Tunneling Protocol") tunnel capable of establishing a point-to-point protocol (PPP) session with the client system 405. The web tunnel 4820 provides a gateway to the routing processor 4812 within the pod 4810, the Internet 465, and a web proxy 4822.

15 The web proxy 4822 can look up subscriber information from the IP address of the client system 405 to determine demographic information such as the subscriber's parental control settings. In this way, the web proxy 4822 can tailor the subscriber's content and user interfaces. The web proxy 4822 can also perform caching functions to store certain URLs ("Uniform Resource Locators") and other electronic content so that the web proxy 4822 can
20 locally deliver information to the client system 405 and avoid the need to access the Internet 465 in the event that data requested by the client system 405 has been cached.

Referring to Fig. 5, a communications system 500 is capable of delivering and exchanging information between a client system 505 and a host system 510 through a communication link 515. Client system 505 typically includes one or more client devices
25 520 and one or more client controllers 525 for controlling the client devices 520. Host system 510 typically includes one or more host devices 535 and one or more host controllers 540 for controlling the host devices 535. The communications link 515 may include communication pathways 550, 555 enabling communications through the one or more delivery networks 560. As shown, the client system 505 may access the Internet 565 through
30 the host system 510.

Examples of each element within the communication system of Fig. 5 are broadly described above with respect to Figs. 1-4. In particular, the client system 505 and the communications link 515 typically have attributes comparable to those described with respect to client systems 105, 205, 305, and 405 and communications links 115, 215, 315, and 415 of Figs. 1-4. Likewise, the host system 510 of Fig. 5 may have attributes comparable to and may illustrate one possible embodiment of the host systems 110, 210, 310, and 410 shown in Figs. 1-4. Fig. 5 describes an aspect of the host system 510, focusing primarily on one particular implementation of IM host complex 590.

The client system 505 includes a client device 520 and a client controller 525. The client controller 525 is generally capable of establishing a connection to the host system 510, including the OSP host complex 580, the IM host complex 590 and/or the Internet 565. In one implementation, the client controller 525 includes an IM application for communicating with servers in the IM host complex 590 using exclusive IM protocols. The client controller 525 also may include applications, such as an OSP client application and/or an Internet browser application, for communicating with elements such as the OSP host complex 580 and the Internet 565.

The host system 510 includes a host device 535 and a host controller 540. The host controller 540 is generally capable of transmitting instructions to any or all of the elements of the host device 535. For example, in one implementation, the host controller 540 includes one or more software applications loaded on one or more elements of the host device 535. In other implementations, as described above, the host controller 540 may include any of several other programs, machines, and devices operating independently or collectively to control the host device 535.

The host system 510 includes a login server 570 capable of enabling communications between client systems 505 and various elements of the host system 510, including elements such as the OSP host complex 580 and IM host complex 590; login server 570 is also capable of authorizing access by the client system 505 and those elements. The login server 570 may implement one or more authorization procedures to enable simultaneous access to one or more of the elements. The OSP host complex 580 and the IM host complex 590 are connected through one or more host complex gateways 585 and one or more IM host complex gateways 595. Each OSP host complex gateway 585 and IM host complex gateway

595 may perform any protocol conversions necessary to enable communication between the OSP host complex 580, the IM host complex 590, and the Internet 565.

To access the IM host complex 590 to begin an instant messaging session, the client system 505 establishes a connection to the login server 570. The login server 570 typically determines whether the particular subscriber is authorized to access the IM host complex 590 by verifying a subscriber identification and password. If the subscriber is authorized to access the IM host complex 590, the login server 570 employs a hashing technique on the subscriber's screen name to identify a particular IM server 5902 for use during the subscriber's session. The login server 570 provides the client system 505 with the IP address of the particular IM server 5902, gives the client system 505 an encrypted key (i.e., a cookie), and breaks the connection. The client system 505 then uses the IP address to establish a connection to the particular IM server 5902 through the communications link 515, and obtains access to that IM server 5902 using the encrypted key. Typically, the client system 505 will be equipped with a winsock API ("Application Programming Interface") that enables the client system 505 to establish an open TCP connection to the IM server 5902.

Once a connection to the IM server 5902 has been established, the client system 505 may directly or indirectly transmit data to and access content from the IM server 5902 and one or more associated domain servers 5904. The IM server 5902 supports the fundamental instant messaging services and the domain servers 5904 may support associated services, such as, for example, administrative matters, directory services, chat and interest groups. The domain servers 5904 can be used to lighten the load placed on the IM server 5902 by assuming responsibility for some of the services within the IM host complex 590. By accessing the IM server 5902 and/or the domain server 5904, a subscriber can use the IM client application to view whether particular subscribers ("buddies") are online, exchange instant messages with particular subscribers, participate in group chat rooms, trade files such as pictures, invitations or documents, find other subscribers with similar interests, get customized news and stock quotes, and search the Web.

In the implementation of Fig. 5, IM server 5902 is directly or indirectly connected to a routing gateway 5906. The routing gateway 5906 facilitates the connection between the IM server 5902 and one or more alert multiplexors 5908. For example, routing gateway 5906 may serve as a link minimization tool or hub to connect several IM servers 5902 to several

alert multiplexors 5908. In general, an alert multiplexor 5908 maintains a record of alerts and subscribers registered to receive the alerts.

Once the client system 505 is connected to the alert multiplexor 5908, a subscriber can register for and/or receive one or more types of alerts. The connection pathway between the client system 505 and the alert multiplexor 5908 is determined by employing a hashing technique at the IM server 5902 to identify the particular alert multiplexor 5908 to be used for the subscriber's session. Once the particular multiplexor 5908 has been identified, the IM server 5902 provides the client system 505 with the IP address of the particular alert multiplexor 5908 and gives the client system 505 an encrypted key (i.e., a cookie) used to gain access to the identified multiplexor 5908. The client system 505 then uses the IP address to connect to the particular alert multiplexor 5908 through the communication link 515 and obtains access to the alert multiplexor 5908 using the encrypted key.

The alert multiplexor 5908 is connected to an alert gate 5910 that, like the IM host complex gateway 595, is capable of performing the necessary protocol conversions to enable communication with the OSP host complex 580. The alert gate 5910 is the interface between the IM host complex 590 and the physical servers, such as servers in the OSP host complex 580, where state changes are occurring. In general, the information regarding state changes will be gathered and used by the IM host complex 590. The alert multiplexor 5908 also may communicate with the OSP host complex 580 through the IM gateway 595, for example, to provide the servers and subscribers of the OSP host complex 580 with certain information gathered from the alert gate 5910.

The alert gate 5910 can detect an alert feed corresponding to a particular type of alert. The alert gate 5910 may include a piece of code (alert receive code) capable of interacting with another piece of code (alert broadcast code) on the physical server where a state change occurs. In general, the alert receive code installed on the alert gate 5910 instructs the alert broadcast code installed on the physical server to send an alert feed to the alert gate 5910 upon the occurrence of a particular state change. Thereafter, upon detecting an alert feed, the alert gate 5910 contacts the alert multiplexor 5908, which in turn, informs the appropriate client system 505 of the detected alert feed.

In the implementation of Fig. 5, the IM host complex 590 also includes a subscriber profile server 5912 connected to a database 5914 for storing large amounts of subscriber

profile data. The subscriber profile server 5912 may be used to enter, retrieve, edit, manipulate, or otherwise process subscriber profile data. In one implementation, a subscriber's profile data includes, for example, the subscriber's buddy list, alert preferences, designated stocks, identified interests, geographic location and other demographic data. The subscriber may enter, edit and/or delete profile data using an installed IM client application on the client system 505 to interact with the subscriber profile server 5912.

Because the subscriber's data is stored in the IM host complex 590, the subscriber does not have to reenter or update such information in the event that the subscriber accesses the IM host complex 590 using a new or different client system 505. Accordingly, when a subscriber accesses the IM host complex 590, the IM server 5902 can instruct the subscriber profile server 5912 to retrieve the subscriber's profile data from the database 5914 and to provide, for example, the subscriber's buddy list to the IM server 5902 and the subscriber's alert preferences to the alert multiplexor 5908. The subscriber profile server 5912 also may communicate with other servers in the OSP host complex 590 to share subscriber profile data with other services. Alternatively, user profile data may be saved locally on the client device 505.

Referring to Fig. 6, a communications system 600 is capable of delivering and exchanging information between a client system 605 and a host system 610 through a communication link 615. Client system 605 typically includes one or more client devices 620 and one or more client controllers 625 for controlling the client devices 620. Host system 610 typically includes one or more host devices 635 and one or more host controllers 640 for controlling the host devices 635. The communications link 615 may include communication pathways 650, 655 enabling communications through the one or more delivery networks 660.

Examples of each element within the communication system of Fig. 6 are broadly described above with respect to Figs. 1-5. In particular, the client system 605 and the communications link 615 typically have attributes comparable to those described with respect to client systems 105, 205, 305, 405 and 505 and communications links 115, 215, 315, 415 and 515 of Figs. 1-5. Likewise, the host system 610 of Fig. 6 may have attributes comparable to and may illustrate one possible embodiment of the host systems 110, 210, 310, 410 and 510 shown in Figs. 1-5. Fig. 6 describes several aspects of one implementation of

the host system 610 in greater detail, focusing primarily on one particular implementation of the login server 670 and IM host complex 690.

The client system 605 includes a client device 620 and a client controller 625. The client controller 625 is generally capable of establishing a connection to the host system 610, including the IM host complex 690. In one implementation, the client controller 625 includes an IM application for communicating with servers in the IM host complex 690 using exclusive IM protocols.

The host system 610 includes a host device 635 and a host controller 640. The host controller 640 is generally capable of transmitting instructions to any or all of the elements of the host device 635. For example, in one implementation, the host controller 640 includes one or more software applications loaded on one or more elements of the host device 635. In other implementations, as described above, the host controller 640 may include any of several other programs, machines, and devices operating independently or collectively to control the host device 635.

The host system 610 includes a login server 670 capable of enabling communications between client systems 605 and various elements of the host system 610, including elements such as the IM host complex 690 and the OSP host complex (580 in Fig. 5); login server 670 is also capable of authorizing access by the client system 605 and those elements. The IM host complex 690 includes an IM server network 6902, a routing gateway 6906, an alert multiplexor network 6908, and one or more alert gates 6910. The IM server network 6902 may include an interconnected network of IM servers and the alert multiplexor network 6908 may include an interconnected network of alert multiplexors. In the implementation of Fig. 6, the IM server network 6902 and the alert multiplexor network 6908 are interconnected by a routing gateway 6906 that serves as a common hub to reduce the number of connections. Each IM server within IM server network 6902 can directly or indirectly communicate and exchange information with one or more of the alert multiplexors in the alert multiplexor network 6908. Each of the alert multiplexors in the alert multiplexor network 6908 may be connected to several alert gates 6910 that receive different types of alerts.

During a session, a subscriber typically will be assigned to one IM server in the IM server network 6902 and to one alert multiplexor in the alert multiplexor network 6908 based on one or more hashing techniques. In one implementation, for example, each IM server in

the IM server network 6902 may be dedicated to serving a particular set of registered subscribers. Because all of the IM servers can communicate with each other, all subscribers can communicate with each other through instant messaging. However, the IM servers and the alert multiplexors are capable of storing subscriber information and other electronic content that may be accessed by the other IM servers and alert multiplexors. Thus, in another implementation, each alert multiplexor in the alert multiplexor network 6908 may be dedicated to storing information about a particular set or subset of alerts. Because all of the alert multiplexors can communicate with each other, all registered subscribers can receive all types of alerts. This networking arrangement enables the load to be distributed among the various servers in the IM host complex 690 while still enabling a subscriber to communicate, share information, or otherwise interact with other subscribers and servers in the IM host complex 690.

Searching

Referring to Fig. 7, an electronic search is performed according to a process 700. The search, which may be a search of the Internet, may be performed, for example, by the systems described above with respect to Figs. 1-6. For instance, process 700 may be performed by one or more of the pods 4804 of Fig. 4. Additionally or alternatively, process 700 may be performed by one or more non-podded servers, such as servers 4806 or farms 4802 of Fig. 4. Process 700 also may be performed by any other hardware component or software component capable of being programmed to receive, process, and send instructions in the manner described.

Process 700 generally includes receiving at least one search term (step 710). The search term then is compared with a list of recommended sites (step 720), previously performed searches (step 730), a hierarchy of category identifiers and terms related to one or more categories (step 740), and an electronic information store that includes content displayed by and/or extracted from different web pages from different web sites (step 750) to determine whether matches exist. Next, a determination is made as to whether a threshold number of matches have been identified between the search term and one or more of the list of recommended sites, the previously conducted searches, the hierarchy of category identifiers, and the electronic information (step 760). An electronic search (e.g., an Internet search using the World Wide Web (WWW)) based on the search term is conducted when less

than a threshold number of matches are identified (step 770). Finally, results that are based on identified matches are displayed (step 780).

Several search terms may be received (step 710) and may be grouped by default as a single string, or may be grouped in other ways. The search terms are typically received from a client system 105, 205, 305, 405, 505, or 605, or from one or more components of the client system, as shown and described in Figs. 1-6. Search terms generally include text defined by letters and/or numbers. However, search terms also may include other searchable content, such as symbols, other alphanumeric characters, and geometric constructs (e.g., arcs); Boolean operators (e.g., AND, OR, ADJ, NOT, NEAR) generally used to define relationships between search terms; parentheses and quotation marks generally used to indicate precision and to group search terms; wild card characters (e.g., ? and *) generally used to represent a portion of a search term; and concept operators (e.g., !) generally used to broaden the search term or phrase to a list of related words related to the search term or phrase in order to search using these related words.

The recommended sites to which the search term is compared (step 720) may include web sites that have been specially designated as recommended sites, web content that is considered proprietary to a web host such as an Internet Server Provider (ISP), or non-proprietary content such as content from an Internet site that has been specially designated to provide content. The recommended sites typically include web site identifiers, such as web site titles, descriptions, and addresses. Web sites may be designated as recommended sites by a human operator, by a process performed by a computer, or otherwise. In any case, criteria used to designate a web site as a recommended site may include, for example, the number of times a site is accessed or the web site content. When matches occur between the search term and one or more of the recommended sites, results are displayed (step 780). The displayed results generally include one or more web site identifiers. An example of a displayed result is shown in Fig. 10b under the heading "Recommended Sites" 1075.

Comparing the search term with previously performed searches to determine whether matches exist (step 730) may include comparing the search term with previously received search terms, such as those stored in an electronic data store (e.g., a memory or a database). The search term also may be compared with the results of previously-performed searches to determine whether matches exist. Based on matches that are determined to exist, results are

displayed (step 780), as shown for example under the heading "Related Hot Searches" 1085 in Fig. 10b. The results generally include a list of search terms for previously-performed searches that share one or more of the received search terms. Selecting one of the results by mouse click or otherwise typically invokes a search process (e.g., process 700) with respect to the chosen result, but may also or alternatively invoke display of the results of previously-performed searches that are retrieved from storage or memory.

Category Searching

Comparing the search term with a hierarchy of category identifiers and/or terms related to one or more categories to determine whether matches exist (step 740) is described in more detail with reference to Fig. 8a, which shows an exemplary process for performing a category search. In the implementation of Fig. 8a, process 740 generally includes receiving at least one search term (step 810), comparing the search term with a hierarchy of category identifiers to determine whether matches exist (step 820), comparing the search term with terms related to one or more categories to determine whether matches exist (step 830), ranking results of the comparisons (step 835), and communicating at least a category identifier based on the matches that are determined to exist within the hierarchy and the terms (step 840).

The search terms received (step 810) generally include the search terms that were received (step 710) for use in performing an electronic search. As such, one or more search terms may be received, and may be grouped together for searching purposes as a single string by default, or may be grouped in other ways.

The hierarchy of category identifiers with which the search terms are compared (step 820) may include identifiers used to represent categories and information relating to those categories. For example, in one implementation, the hierarchy of category identifiers may include a hierarchy of category names, where groups of the category names are linked together in a hierarchical relationship. In this instance, names in the hierarchy represent categories, the names of which are linked together using sub-categories. The hierarchy of category identifiers also may include other related information, such as a list of web sites that are related to the category by name, description, or otherwise.

Referring to Fig. 8b, an exemplary screen shot 850 illustrates an example of a hierarchy of category identifiers 855. In this instance, the hierarchy of category identifiers 855 appears as a linked list of category names that are affiliated with specific categories. The hierarchical relationship among category identifiers and other related information typically is ordered with broad category names and information followed by more narrow names and information. Other forms and data contents also may be used to express a hierarchy of category identifiers. For instance, the category identifier may additionally or alternatively include other information representing categories therein, such as text, alphanumeric characters, symbols and combinations thereof. In one implementation, some or all of the hierarchy of category identifiers may be arranged by and/or received from a third party listing service (e.g., Open Directory Project).

In comparing the search terms (step 820), matches are typically determined to exist when a received search term matches one or more of the identifiers within the hierarchy of category identifiers. When several received search terms are grouped as a single string for searching purposes, the comparison includes comparing the single string of search terms with the hierarchy of category identifiers to determine whether matching strings exist.

Comparing the search term with terms related to one or more categories to determine whether matches exist (step 830) may include using information related to categories, such as a name of a web site corresponding to a category, a description of the web site, or other related terms. When several search terms are received and grouped as a single string, the comparison (step 830) may include comparing the single string of search terms with the terms related to one or more categories to determine whether matches exist. To improve searchability of terms provided by third party listing services (e.g., Open Directory Project), the comparison may include converting received or related terms to a predesignated searchable format, e.g., by indexing and cataloguing the terms.

Ranking the results (step 835) generally includes ranking the results of comparisons performed in either, both, or the combination of steps 820 and 830. The category identifiers may be ranked based on at least one of a number of matches that are determined to exist, the relative locations of matches, and the relative types of matches. For instance, the potential relevance of a matching category is generally deemed to increase as the number of identified matches increase. Furthermore, the potential relevance, and hence the rank, of a matching

category is deemed to change based on the existence and frequency of matches that occur within different types of information, such as the hierarchy of category identifiers (step 820) and the terms related to one or more categories (step 830). For instance, the results may be ranked based on the existence and number of matches between a search term and the hierarchy of category identifiers, or within the terms related to one or more categories.

Ranking of matching categories and corresponding category identifiers also may be based on the relative location of the matches within the hierarchy of category identifiers. For example, a match occurring in a category identifier that represents a narrow category may be ranked higher than a match occurring within a category identifier that represents a broader category, or vice versa. Further, category identifiers that include matches occurring within the hierarchy of category identifiers are generally ranked higher than category identifiers that include matches that occur within the terms related to one or more categories. Ranking the category identifiers based on the type of the match also may include ranking the category identifiers based on whether the matches occur within at least one of the terms related to one or more categories and the hierarchy of category identifiers. When category identifiers include matches occurring within more than one type, those identifiers are ranked higher than category identifiers that contain matches occurring within only one of the types. For example, a category having matches occurring within both the hierarchy of category identifiers and the terms related to one or more categories is typically ranked higher than a category that includes matches occurring within only one of the hierarchy of category identifiers and the terms related to one or more categories.

Communicating at least a category identifier (step 840) generally includes communicating information revealing matches that are determined to exist within the hierarchy and the related terms. The results communicated generally include at least a category identifier, and are provided for use in a displaying process, such as display step 780 of Fig. 7, for eventual display to a user of a client system.

Referring to Fig. 8b, an example of results communicated (step 840) is displayed under the heading, "Matching Categories." The matching categories of Fig. 8b include several hierarchies of category identifiers, with each hierarchy of category identifiers being ranked in terms of relevance to the proffered search term. The hierarchy of category identifiers shown by Fig. 8b is a listing of category identifiers. The listing starts with an

identifier 8551 for a broad category and descends to an identifier for a more narrow category, with the last category identifier 8552 being the final matching category name. A hierarchy that includes a match within the final category name is generally ranked higher (step 830) than a hierarchy that includes a match within a category name other than the final category name within the hierarchy of category identifiers. In one implementation, selecting one of the categories using a mouse or otherwise will reveal another screen shot 860, as shown, e.g., in Fig. 8c. Each category may include a listing of sub-categories 865 and web sites 875 within those categories. For instance, the listing for a web site within a category may include the title of the web site, a description of the web site, and an address for the web site.

Matching Sites Search

Referring to Fig. 9a, another aspect 750 of the search process 700 shown by Fig. 7 is described for performing a search to identify web sites that relate to a search term. In this search process 750, the search term is compared against text or other searchable content displayed or extracted from the actual web site(s). Process 750 of Figs. 7 and 9a generally includes receiving at least one search term (step 910), comparing the search term with electronic information within at least one electronic information store to determine whether matches exist (step 920), ranking results of the comparison (step 925), and communicating results based on the matches that are determined to exist (step 930).

The search terms received (step 910) generally include the search terms that were received (step 710) for use in performing an electronic search. As such, one or more search terms may be received, and may be grouped together for searching purposes as a single string by default, or may be grouped in other ways.

The received search terms may be compared (step 920) to electronic information within at least one electronic information store to determine whether matches exist. The electronic information may include, for example, text or other searchable content displayed by and/or extracted from web pages from different web sites. When several search terms are received, the comparison (step 920) may include comparing the single string of search terms with the electronic information within the electronic information store to determine whether matches exist. The electronic information may include partial or full text displayed by

different web pages from different web sites (e.g., an introductory or home page), titles, descriptions, and addresses of web sites.

Ranking the results (step 925) generally includes ranking search results based on an algorithm that takes into account various aspects of the results achieved. For example, the identifiers for the several web sites may be ranked based on a number of the matches that are determined to exist between the search term and the electronic information corresponding to the web sites. Ranking the identifiers for the several web sites also may be based on whether matches occur within one or more of the text, the title, the description, and the addresses of the web site. For instance, identifiers with more than one of the title, description, text, and web address that match a search term are generally ranked higher than identifiers with only one of the title, description, text, and web address that match the same search term.

In addition, the ranking also may be based on which of these forms of electronic information are matched and where the matches occur. For example, identifiers with titles that match a search term may be ranked higher than identifiers with descriptions that match the same search term, which may be ranked higher than identifiers with web addresses that match the same search term, which may be ranked higher than identifiers with text that matches the same search term.

Communicating results (step 930) may be based on matches that are determined to exist from the comparison (step 920). For instance, the search results communicated (step 930) may be provided for use in a displaying process, such as displaying step 780 of Fig. 7, for eventual display to a user of, e.g., a client system. The results communicated typically include an identifier for each matching web site, such as a title, a description, address information, text, characters, symbols, or combinations thereof used to identify or describe a web site. For example, Fig. 9e shows an exemplary display 990 of identifiers 932.

Filtering Search Results

Referring also to Fig. 9b, comparing the search term with electronic information within at least an electronic information store (step 920) may include classifying the search term among at least first and second categories (step 922), comparing the search term to first electronic information within a first electronic information store to determine whether matches exist when the search term is classified within the first category (step 924), and

comparing the search term to at least second electronic information within at least a second electronic information store to determine whether matches exist when the search term is classified within the second category (step 926). As shown in Fig. 9a, the process 920 of comparing shown by Fig. 9b may be preceded by receiving at least one search term (step 910 of Fig. 9a) and followed by ranking and communicating a result based on the matches that are determined to exist (steps 925 and 930 of Fig. 9a). A more detailed description for steps 910, 925, and 930 is provided above with reference to Fig. 9a; a more detailed description of steps 922, 924, and 926 is provided below.

Classifying the search term (step 922) generally includes classifying the received search term among one or more categories, with a first category and a second category being described and shown for illustrative purposes. If several search terms are grouped as a single string, the search terms may be collectively classified as a single string based on the grouping of the search terms, or they may be classified individually based on each individual search term.

Comparing the search terms (step 924) generally includes comparing the search term to first electronic information within a first electronic information store when the search term is classified within the first category. By contrast, comparing the search term (step 926) generally includes comparing the search term to the second electronic information within the second electronic information store to determine whether matches exist when the search term is classified within the second category. However, comparing the search term (step 926) may also include comparing the search term to the first electronic information within the first information store such that matching results from both electronic information stores may result from the comparison (step 926). In this instance, search terms are compared to a first set of data (step 924), and compared to a second set of data that includes the first set of data and other data (step 926).

Referring to Fig. 9c, a system that searches and stores searchable content includes first and second electronic information stores 992 and 994 which store electronic information received or derived from different sources which may have different classifications. The system may further include additional electronic information stores as illustrated by item 996, and generally may include a search engine 998 for comparing received search terms with the content within either or both information stores to determine whether matches exist.

5 The first electronic information store 992 and the second electronic information store 994 may be a part of a single storage device or several separate storage devices, examples of which include a magnetic disk (e.g., an internal hard disk and removable disk); a magneto-optical or optical disk; and a CD-ROM. The first electronic information store 992 and the second electronic information store 994 also or alternatively may be a part of a single volatile or non-volatile memory device or several separate non-volatile memory devices, examples of which include semiconductor memory devices such as RAM, ROM, PROM, EPROM, EEPROM, and flash memory devices. When stored on separate devices, the first electronic information store may be located on a first server and the second electronic information
10 device may be located on a second server that differs from the first server.

15 The first and second electronic information stores 992 and 994 each may include partial or full text or other searchable content displayed by one or more different web pages from one or more different web sites, and may include identifiers for those web sites, such as titles, descriptions, and addresses. The first and second electronic information stores 992 and 994 are typically populated by automatically scanning and storing the text and/or other searchable content of a web site that has been accessed a threshold number of times by members of a web host as described with respect to steps 940, 945, and 950 of Fig. 9d, or that has been identified by a listing service (but not accessed the threshold number of times by members of the web host) as described with respect to steps 960, 970, and 980 of Fig. 9d.
20 However, either of stores 992 and 994 may be populated in other ways. In either case, the first and second electronic information stores 992 and 994 store searchable content corresponding to the contents scanned from web pages, whether identified through access activities, list service identification, or in other ways.

25 Each electronic information store may contain content that has been classified and stored based on a specified type or types of classification criteria. For instance, the first electronic information store 992 may include content classified as non-offensive and the second electronic information store 994 may include content classified as offensive. Other types of content classification criteria may be implemented in addition to or separate from criteria based on offensive and non-offensive classifications. Other criteria that may be used,
30 for example, include medical and non-medical, legal and non-legal, and sports and non-sports.

In one implementation, the first electronic information includes contents relating to non-offensive web sites, and the second electronic information includes contents relating to offensive web sites. Example of non-offensive web sites may include web sites that do not include pornographic, violent, racist, or hate-related content. By contrast, examples of offensive web sites may include web sites that include pornographic, violent, racist, or hate-related content.

The following describes an example applying the described search methods of Fig. 9b to this implementation. A user of a client system enters a search term (step 910). The search term is classified as either being offensive or non-offensive (step 922). If the term is classified as being non-offensive, then only the contents of the first electronic information store are searched (step 924) and results from the search are communicated for display to the user (step 930). In this example, the first electronic information store only contains contents that previously have been classified as non-offensive. If the search term entered by the user is classified as being offensive, the contents of either the second electronic information store or both the first and second electronic information stores are searched (step 926) and the results are communicated for display to the user (step 930).

The described filtering of results between offensive content and non-offensive content based on the classification of the search term may allow a web host to implement a parental type of control in determining what search results are displayed to the user. Because the offensive and non-offensive contents are stored in different electronic information stores, the ability to restrict access is enhanced. For instance, parental control can be exercised by blocking the access of a user to one or more electronic information stores. Other forms of data filtering also are enabled through this process and related techniques.

Referring to Fig. 9d, the electronic information within the electronic information store may be populated by various methods. For instance, process 750 of Figs. 7 and 9a also may include identifying web sites and/or web pages accessed by members of a web host (step 940), automatically scanning the text of a web site when the web site is accessed by a member of a web host (step 945), storing text or other searchable content from within the web site that was automatically scanned for comparison against search terms that were received (step 950), identifying web sites provided by a listing service (step 960), determining whether text or other searchable content for web sites identified by the listing

service were stored previously (step 970), and automatically scanning and storing text or other searchable content from within web sites that were determined not to be stored previously (step 980). The relative order of steps within Figs. 9a and 9d should not be construed to imply order among the steps described by those respective figures.

5 The access activity of members of a web host may be monitored to enable web sites that have been accessed to be identified for scanning and storage in preparation for future electronic searches (step 940).

10 Automatically scanning (step 945) typically includes automatically scanning the text of a web site when the web site is accessed by a member or a configurable threshold number of members of the web host. Automatically scanning also may include scanning the full text of the web site, scanning text included on an introductory page, and scanning full text included on an introductory page. Scanning generally includes character or image recognition techniques, but may include other methods of capturing and conversion of information displayed by accessed web pages or web sites to searchable form.

15 Storing text (step 950) generally includes storing the text or other searchable content of the web site that was automatically scanned for future comparison against search terms. The text may be stored in an electronic information store such as those described above, which may be embodied, for example, by cache memory.

20 In one implementation, prior to scanning (step 945) and storing (step 950), populating the electronic information store may further include identifying the web site being accessed by a member of the web host (step 940) and determining whether the text of the web site was previously stored. In this implementation, automatically scanning and storing will occur when the text of the web site is determined not to have been previously stored. Otherwise, the scanning and storing may be skipped.

25 Determining whether the text of the web site has been stored may be accomplished using various methods. For example, determining whether the text of the web site has been stored may be based on a web site address that corresponds to the web site being accessed by the member of the web host. Additionally or alternatively, determining whether the text has been stored may be based on the text of the web page or on the web site itself.

30 Another method for populating the electronic information within the electronic information store includes identifying web sites provided by a listing service (step 960),

determining whether the text of the web sites provided by the listing service were scanned and stored previously (step 970), and automatically scanning and storing the text of the web sites determined not to be stored previously (step 980).

Identifying web sites provided by a listing service (step 960) may occur on a periodic basis (e.g., daily, weekly, monthly), based on a triggering event (e.g., receipt of listing service information), or otherwise. The listing service generally includes a third party service such as that provided by the Open Directory Project (ODP).

Determining whether web site content has been previously scanned and stored (step 970) may include searching memory or storage contents for content or identifiers corresponding to the web page or web host, either through a search of a table of contents for the memory or storage, or through a search of the memory or storage itself.

Automatically scanning and storing (step 980) also may include automatically scanning and storing the full text of web sites provided by the listing service.

The process for searching different data stores may further include automatically scanning contents of a web site when the web site is accessed by a member of a web host, classifying the contents of the web site among at least one of the first electronic information within the first electronic information store and the second electronic information within the second electronic information store, storing the contents of the web site as part of the first electronic information when the contents are classified among the first electronic information and storing the contents as part of the second electronic information when the contents of the web site are classified among the second electronic information.

Displaying Web Site Search Results

Referring to Fig. 10a, a process 780 for displaying web site search results generally includes receiving at least one search term (step 710). The search term is compared with first electronic information within a first electronic information store including content provided by an internal source to determine whether matches exist (step 1020). The search term also is compared with second electronic information within a second electronic information store including content provided by an external source to determine whether matches exist (step 1030). Results based on the matches that are determined to exist with the first electronic information and the second electronic information are displayed, with the results combined in

a single list of results including the matches that are determined to exist with the first electronic information and the second electronic information (step 1040).

Typically, at least one search term is received (step 710). However, several search terms may be received and may be grouped by default as a single string, or may be grouped in other ways. The search terms may be received from any type of source (e.g., a user of a client system, a search engine, a component of a process for searching the Internet).

Comparing with first electronic information (step 1020) typically includes comparing the search term to first electronic information within a first electronic information store to determine whether matches exist. The first electronic information may include content provided and/or maintained by a web searching host, and content for which access is provided only to selected members by the web searching host. The web searching host may be an Internet service provider or some other content maintaining and providing service. The content may include content that is proprietary to the web searching host and content that is proprietary to another entity, but that is made accessible only to members of the web searching host.

Comparing with second electronic information (step 1030) typically includes comparing the search term to second electronic information within a second electronic information store to determine whether matches exist. The second electronic information may include content provided and/or maintained by a source external to the web searching host. One example of second electronic information includes content that is available to both members and non-members of a web searching host, such as content available to any member of the public on the World Wide Web. The content may include content that is non-proprietary to the web host as well as content that is proprietary to another entity, but that is available to others as well as to members of the web host.

For instance, steps 1020 and 1030 may correspond to searching processes described with respect to one or more of steps 720, 730, 740, and 750, where the web searching host searches its own content as well as externally provided and maintained content for matches with any or all of recommended sites, previously performed searches, category identifiers, and electronic information such as text from the web pages. An example of a web searching host includes America Online (AOL), which maintains web-accessible contents and which

enables searching of those web-accessible contents and other non-AOL maintained contents, with display of amalgamated search results.

Results may be displayed (step 1040) based on the matches that are determined to exist with the first electronic information and the second electronic information. The results may be combined in a single list of results. Displaying (step 1040) may include displaying results such that the source of the results obtained from the external source or the web searching host is transparent to a user viewing the single list of results. Furthermore, the single list of results may be displayed in a ranked list of results. The ranking may be in descending order of relevance from results that are most relevant to the received search term to results that are least relevant to the received search term. Each result may be assigned a relevance weighting based on the numerous factors that may be considered by a ranking algorithm. Some of the factors used by the ranking algorithm may include the number of shared words between the search term and the results, and the identification of the component of a single result (e.g., title of the web site, description of the web site, address of the web site, text of the web site) in which the shared terms occur. Additionally or alternatively, the results may be ranked according to whether the match occurs between the search term and the internal source or between the search term and the external source.

For example, as shown in Fig. 9e under "Matching Sites", a single ranked list of results is displayed so that the source of any one listed result is transparent to a viewer of the results. Similarly, as shown in Fig. 10c under "Matching Web Pages", a single ranked list of results is displayed so that the source of any one listed result is transparent to a viewer of the results. A viewer of the results is unaware of the proprietary or non-proprietary nature of any of the results.

In one implementation, a process for displaying web site search results that are produced from searching multiple electronic information stores generally includes sending the search term to a third party search service for use in comparing the search term to at least second electronic information within a second electronic information store, receiving the results from the third party search service, combining the first results and the second results, and displaying the combined first results and second results as a single list of results, with the results including at least one web site identifier.

In this implementation, the first electronic information may include proprietary information and the second electronic information may include electronic information that is non-proprietary to a provider of the first electronic information. The second electronic information may be maintained by a third party search service and may include information that is proprietary to the third party search service. The first electronic information within the first electronic information store may be maintained by an Internet service provider.

In addition, the systems, methods, and techniques described here may be implemented in digital electronic circuitry, or in computer hardware, firmware, software, or in combinations of them. Apparatus embodying these techniques may include appropriate input and output components, a computer processor, and a computer program product tangibly embodied in a machine-readable storage component for execution by a programmable processor. A process embodying these techniques may be performed by a programmable processor executing a program of instructions to perform desired functions by operating on input data and generating appropriate output. The techniques may advantageously be implemented in one or more computer programs that are executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input component, and at least one output component. Each computer program may be implemented in a high-level procedural or object-oriented programming language, or in assembly or machine language if desired; and in any case, the language may be a compiled or interpreted language. Suitable processors include, by way of example, both general and special purpose microprocessors. Generally, a processor will receive instructions and data from a read-only memory and/or a random access memory. Storage components suitable for tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory components, such as Erasable Programmable Read-Only Memory (EPROM), Electrically Erasable Programmable Read-Only Memory (EEPROM), and flash memory components; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and Compact Disc Read-Only Memory (CD-ROM disks). Any of the foregoing may be supplemented by, or incorporated in, specially-designed ASICs (application-specific integrated circuits).

Accordingly, other embodiments are within the scope of the following claims.